WHO DID ROLLO SIGN THE TREATY OF SAINT-CLAIR-SUR-EPTE WITH?

+

...THE TREATY OF SAINT-CLAIR-SUR-EPTE BETWEEN KING CHARLES III OF WEST FRANCIA AND THE FAMED VIKING RULER ROLLO...

RYAN QA

RYAN RICE | RYANRICE@STANFORD.EDU

WHO DID ROLLO SIGN THE TREATY OF SAINT-CLAIR-SUR-EPTE WITH?

=

KING CHARLES III OF WEST FRANCIA

THE PROBLEM

OPEN DOMAIN QUESTION ANSWERING IS A FIELD OF NLP THAT REQUIRES SYSTEMS TO ANSWER QUESTIONS OF ANY TOPIC AND CONSEQUENTLY DEVELOP A MORE COMPLETE UNDERSTANDING OF LANGUAGE.

GOOGLE'S BERT IS A PRE-TRAINED LANGUAGE MODEL THAT HAS HAD SUCCESS IN A VARIETY OF NLP TASKS, AND EXTENSIONS OF BERT ARE CURRENTLY STATE-OF-THE-ART FOR QUESTION ANSWERING.

THE DATA

THE SQUAD 2.0 DATASET WAS USED TO TRAIN THE MODEL.

IT CONSISTS OF CONTEXT-QUESTION PAIRS AS INPUTS AND EITHER CORRESPONDING ANSWER SPANS FROM THE CONTEXT OR NO ANSWER AS OUTPUTS.

THE DATA SPLIT IS AS FOLLOWS: 129941 TRAINING EXAMPLES, 6078 DEV EXAMPLES, AND 5915 TEST EXAMPLES.

THE MODEL

THE MODEL IS A WEIGHTED ENSEMBLE OF NETWORKS TRAINED BY FINE-TUNING THE BASE BERT MODEL FOR QUESTION ANSWERING [1].

TWO TYPES OF ARCHITECTURES AND THREE MAX SEQUENCE LENGTHS WERE TESTED: VANILLA/ATTENTION AND 128/256/384.

THE VANILLA MODELS UTILIZE A SINGLE LINEAR LAYER STACKED ON BERT FOR OUTPUT WHILE THE ATTENTION MODELS UTILIZE MULTIPLICATIVE ATTENTION FOR GENERATING ANSWER SPANS.

THE RESULTS

WHILE A VARIETY OF ENSEMBLE COMBINATIONS AND WEIGHTS WERE TRIED, THE ENSEMBLE USING THE 128V, 128A, 256V AND 384V MODELS (WITH THE 256V AND 384V MODELS WEIGHTED TWICE THE OTHERS) PERFORMED THE BEST.

INDIVIDUALLY, THE 256V MODEL HAD THE HIGHEST PERFORMANCE.

OVERALL, THE MODEL ACHIEVED 75.716 EM AND 78.338 F1 OUTPERFORMING THE BASELINE BY +19.725 EM AND +19.047 F1.

ANALYSIS

BECAUSE ROUGHLY HALF OF THE DATASET HAS UNANSWERABLE QUESTIONS, MODELS TEND TO BIAS TOWARD INCORRECTLY MARKING QUESTIONS AS UNANSWERABLE.

THE WEIGHTED ENSEBMLE ALLOWS THE MODEL TO OVERCOME THIS BIAS BY INCORPORATING THE 5 MOST LIKELY ANSWERS FROM A VARIETY OF ARCHITECTURES THAT SUCCEED AT DIFFERENT TYPES OF QUESTIONS. FOR THE EXAMPLE ABOVE, THE 384V SINGLE MODEL – THE MODEL WITH THE LARGEST CONTEXT – INCORRECTLY DOES NOT ANSWER THE QUESTION; HOWEVER, THE ENSEMBLE GETS THE QUESTION CORRECT BECAUSE OF ITS OTHER COMPONENTS. WHILE NOT PERFECT, THE WEIGHTED ENSEMBLE APPROACH DOES BOOST PERFORMANCE.

CONCLUSION

MAXIMUM SEQUENCE LENGTH APPEARS TO HAVE THE GREATEST IMPACT ON THE PERFORMANCE OF THE MODEL. THAT BEING SAID ENSEMBLE MODELS CAN STILL BENEFIT FROM SINGLE MODELS USING A VARIETY OF SEQUENCE LENGTHS AS THE MODELS SUCCEED ON DIFFERENT TYPES OF QUESTIONS. FINALLY, GIVEN MORE TIME AND RESOURCES, THE EFFECTS OF USING THE LARGE BERT MODEL (AS OPPOSED TO THE BASE MODEL USED HERE) WOULD BE EXPLORED.

[1] JACOB DEVLIN ET AL.
BERT: PRE-TRAINING OF DEEP BIDIRECTIONAL TRANSFORMERS FOR
LANGUAGE UNDERSTANDING. 2018.